DOCUMENT RESUME

ED 219 458 TM 820 493

TITLE	Reading Test for New York State Elementary Schools, Grade 6-Form E. Manual for Administrators and Teachers.
INSTITUTION	New York State Education Dept., Albany. Div. of Educational Testing.
PUB DATE	82
NOTE	51p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Difficulty Level; Grade 6; Intermediate Grades; Latent Trait Theory; Readability; *Reading Comprehension; Reading Diagnosis; Reading Improvement; *Reading Tests; Remedial Instruction; *State Programs; *Testing Programs; Test Validity
IDENTIFIERS	*Degrees of Reading Power; Pupil Evaluation Program; *Test Manuals

ABSTRACT

The Reading Test for New York State Elementary Schools is designed to determine the most difficult prose text a student can read with comprehension. Test uses described include evaluating the student's current level of achievement; determining reading difficulty levels for instruction; measuring growth of comprehension; determining statewide comprehension trends; and indicating remediation needs for success on the Regents Competency Test in Reading. The Reading Test consists of 11 nonfiction prose passages on selected topics organized into 77 test items formed by passage deletions requiring responses. The Degrees of Reading Power (DRP) test items' relationship to cloze procedures are shown. The test materials, preparation procedures, specific directions for administering the test, and recommendations for handscoring are discussed. In interpreting the results, the role of the State Reference Point cutoff for remediation is presented. Discussions of DRP methodology in instructional use, prose difficulty and readability scales: the validity and reliability of the test; and problems with grade-equivalent and grade-level readability scores cover the technical characteristics of the reading test. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219458

READING TEST

FOR

NEW YORK STATE ELEMENTARY SCHOOLS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Grade 6 — Form E

Manual for Administrators and Teachers

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM C

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Trombly

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM820153

Copyright 1982

The University of the State of New York
The State Education Department
Albany, New York 12234

THE UNIVERSITY OF THE STATE OF NEW YORK

Regents of The University *(with years when terms expire)*

1988 WILLARD A. GENRICH, *Chancellor*, LL.B., L.H.D., LL.D., Litt.D., D.C.S., D.C.L. ---------- Buffalo

1988 J. EDWARD MEYER, *Vice Chancellor*, B.A., LL.B. ---------- Chappaqua

1986 KENNETH B. CLARK, A.B., M.S., Ph.D., LL.D., L.H.D., D.Sc. ---------- Hastings on Hudson

1983 HAROLD E. NEWCOMB, B.A., LL.D. ---------- Owego

1982 EMLYN I. GRIFFITH, A.B., J.D. ---------- Rome

1983 MARY ALICE KENDALL, B.S. ---------- Rochester

1984 JORGE L. BATISTA, B.A., J.D., LL.D. ---------- Bronx

1986 LAURA BRADLEY CHODOS, B.A., M.A. ---------- Vischer Ferry

1987 MARTIN C. BARELL, B.A., I.A., LL.B. ---------- Kings Point

1984 LOUISE P. MATTEONI, B.A., M.A., Ph.D. ---------- Bayside

1987 R. CARLOS CARBALLADA, B.S., L.H.D. ---------- Rochester

1988 FLOYD S. LINTON, A.B., M.A., M.P.A., D.C.L. ---------- Miller Place

1988 SALVATORE J. SCLAFANI, B.S., M.D. ---------- Staten Island

1982 MIMI LIEBER, B.A., M.A. ---------- New York

1985 SHIRLEY C. BROWN, B.A., M.A., Ph.D. ---------- Albany

President of The University and Commissioner of Education
GORDON M. AMBACH

Executive Deputy Commissioner of Education
ROBERT J. MAURER

FOREWORD

The Degrees of Reading Power (DRP) methodology is part of the Pupil Evaluation Program for several reasons. First, the DRP provides teachers with an effective tool for matching instructional materials with student ability. Second, since DRP tests can validly be used to measure individual gain, they can be used to evaluate the effectiveness of reading programs including Title I and PSEN programs. Finally, since DRP test methodology is the basis for the Preliminary and Regents Competency Tests in Reading, the use of DRP tests in the intermediate grades is valuable for providing students with appreciable experience using the DRP test format, as well as for furnishing concrete information concerning the growth required to meet the diploma standard in reading.

The exact composition of a DRP test depends on the purpose or function of the program for which the test was designed. For example, when DRP methodology is used for the Regents Competency Test, the test passages are concentrated around the graduation standard established by the Board of Regents. When it is used for the sixth-grade Pupil Evaluation Program and the Preliminary Competency Test, a broad range of passages is used in order to identify students in need of various types of remedial activity.

CONTENTS

	Page
Foreword	i
Description of the Test	
Purpose of the Test	1
General Features of the Test	1
DRP vs. Cloze	1
DRP vs. Multiple Choice Cloze	2
Administering and Scoring the Test	
The Test Materials	2
Special Considerations for Assuring Optimal Student Performance	3
Test Administration Time	3
Preparations for Testing	4
Detailed Directions for Administering the Test	4
Scoring the Test	7
Interpreting Test Results	
The State Reference Point and Remediation Requirements	8
Notifying Students and Parents About Test Results and Placement Decisions	10
Required Records	10
Recommendations: Students Scoring Above the State Reference Point	10
Recommendations: Students Scoring Below the State Reference Point	11
Using the Test for Instructional Purposes	11
Prose Difficulty or Readability	12
The Measurement of Readability	13
Reading Ability and Prose Difficulty	13
Converting Raw Scores to Reading Ability Levels	17
Using the Test as a Diagnostic Instrument	20
Organizing Reading Programs	22
Using the Test as a Basis for Terminating Remediation	26
Technical Characteristics	
Validity and Reliability of the Test	26
Grade-Equivalent and Grade-Level Readability Scores	32

Appendixes

I List of Correct Answers 34
II Sample Letters to Parents 36
III Derivation and Validity of DRP Units 39
IV Request for Readability Analysis 43

DESCRIPTION OF THE TEST

Purpose of the Test

The Reading Test for New York State Elementary Schools is designed to measure the most difficult prose text a student can read with comprehension. Scores obtained on the test can be used for: (1) evaluating students' current level of achievement in reading; (2) determining the most difficult prose text a student can profitably use in instruction and in independent reading; (3) measuring growth in the ability to read with comprehension; (4) determining statewide trends in students' ability to read with comprehension; (5) indicating the extent of compensatory or remedial help, if any, that a student may need in order to achieve success on the Regents Competency Test in Reading; and (6) providing the basis for the termination of remediation. The test is administered to sixth-grade students as part of the Pupil Evaluation Program.

General Features of the Test

The Reading Test for New York State Elementary Schools consists of 11 nonfiction prose passages on randomly selected topics. Each passage contains about 300 words. The passages are arranged in order of difficulty, beginning with very easy material and progressing to very difficult material. Test items are formed by the deletion of seven words in each passage, and each deleted word is indicated by an underlined blank space. For each deletion, five single-word response options are provided. Students are required to read the text, select the word that most appropriately fits the blank, and mark their answers on separate answer sheets. There are a total of 77 items on the test, which requires about 75 minutes of testing time.

DRP vs. Cloze

Although DRP test items are formed by the deletion of words in the passages, the DRP is not a cloze test. A test is correctly classified as cloze only if it exhibits the following characteristics:

1) The words to be deleted are selected by a mechanical procedure such as leaving out every 5th or 10th word or by selecting words through use of a table of random numbers. Words may

be deleted throughout the length of the entire article or deletions may be restricted to samples of lines from the total passage, provided the line samples are chosen by some random or "every nth" procedure. In the DRP, words are not deleted by a mechanical process.

2) It is a generative task. Students must guess the word that belongs in the blank. On the DRP, response options are provided.

3) Success or failure in guessing one word may influence subsequent success or failure in guessing other words. In the DRP, items must be independent.

DRP vs. Multiple Choice Cloze

There are also numerous factors that distinguish DRP from a multiple choice cloze.

1) Unlike multiple choice cloze, DRP completion questions are constructed when all the content information required to answer the question has been provided. The DRP requires that each item must be answerable on the basis of information presented.

2) On the DRP, *each* response option, when inserted in the blank, must make a grammatically correct and semantically plausible sentence when the sentence frame is considered in isolation.

ADMINISTERING AND SCORING THE TEST

The Test Materials

The test materials provided by the Department for the Reading Test for New York State Elementary Schools, Grade 6, Form E, include: (1) a test booklet, (2) an NCS answer sheet (hand and machine scorable), (3) a handscoring overlay, and (4) a manual for administrators and teachers.

Hand scoring is recommended so that teachers may become familiar with a student's performance. It will also allow them to make informed inferences about carelessness and guessing, which may affect decisions concerning who should be assigned to a particular reading group and/or who should be assigned what reading materials.

For machine processing, several grids containing spaces for recording various types of student identification information have been included on the answer sheet. If the answer sheets are to be machine scored, it is essential that uniform written instructions about the completion of these grids be developed by the school and provided to all examiners administering the test. Such instructions should be based on careful consideration of the types of student and score information the school needs, as well as on the processing requirements of the scoring center which the school is using. Note that the NCS answer sheet is the only one which is supplied by the Department.

Special Considerations for Assuring Optimal Student Performance

If the test results are to provide an accurate measure of each student's reading ability, the directions for administering the test should be followed carefully. The person administering the test should review the directions beforehand and become thoroughly familiar with them. Since the purpose of the test is to obtain the most accurate estimate possible of a student's reading ability, it is essential that students be given an opportunity for fair assessment.

Of course, students should not be given help in interpreting the test questions. They should be advised to answer the questions according to their best judgment. However, the examiner should give students all the assistance required in such mechanics as filling out the heading of the answer sheet and recording the answers.

Test Administration Time

The Reading Test for New York State Elementary Schools is untimed, and students are to be given as much time as they need to answer as many questions as they can. As a general rule, most students will need about 1 hour of working time; in addition, approximately 15 minutes will be necessary for the examiner to provide instructions to the students. Therefore, schools should schedule at least 1½ hours for the test administration. However, some students may need more working time for the test, and schools should make provisions to accommodate such students.

Preparations for Testing

Although special arrangements may be needed for administering the test to large groups of students, these arrangements should interfere as little as possible with the normal school routine. A number of points should be taken into consideration when making arrangements for the testing session.

1. Selecting the Testing Room. If a room other than the regular classroom is to be used, attention should be given to lighting, ventilation, and the absence of noise and other distractions.
2. Scheduling the Test. The test should be scheduled, if possible, on Tuesday, Wednesday, or Thursday. The test should not be scheduled immediately after a period of strenuous physical exercise.
3. Orientation of Students. Any announcements informing students about the test before the test administration should be made in such a way as to increase the students' interest in the test and at the same time not to cause them to become overly anxious and tense. They should be told that no special preparation is necessary and that the test measures reading skills which they use every day, in and out of school.
4. Assembling the Test Materials. The materials should be assembled at least 1 day before the test administration. The following materials are needed for this test:

   For each student:
   - Test booklet
   - NCS answer sheet
   - No. 2½ or softer lead pencil

   For the examiner:
   - Directions for administering and scoring
   - Test booklet (for demonstration purposes)
   - NCS answer sheet (for demonstration purposes)
   - Instructions for completing the student identification grids on the NCS answer sheet (provided by the school if the answer sheets are to be machine scored)
   - Extra answer sheets and pencils

Detailed Directions for Administering the Test

Make sure that sufficient quantities of all test materials are at hand. Each student should have a No. 2½ or softer lead pencil and the desks should be cleared of books and papers.

When the students are ready to begin the test, say:

> The test you are about to take is a new type of reading test. It is not a timed test. You will be given as much time as you need.
>
> I will now give out the test booklets.

Distribute one test booklet, face up, to each student and say:

> Please leave the test booklet face up on your desk. Do not open the booklet until I am ready to read through the directions with you.

While distributing the answer sheets say:

> Do not write on the answer sheets until I tell you what to do.

After the answer sheets have been distributed, give directions for completing the heading. **Be sure that students clearly indicate on the answer sheet the form of the test they will be taking.** The form of the test may be found on the cover of the test booklet. If the answer sheets are to be machine scored, provide directions for marking any machine-readable name or number grids.

After students have provided the information requested on the answer sheets say:

> Look at the cover of your test booklet. Read the section called "To the Student" to yourself while I read it aloud.
>
> "To the Student
>
> This is a test to find out how well you read. Do your best to read the passages and answer the questions. Work carefully and do not rush. You will be given as much time as you need."

Now, open your test booklet to page 3. Read the directions to yourself while I read them aloud.

"Directions:

This test contains several reading passages. In each passage some words are missing. Wherever a word is missing, there is a blank line with a number on it. Next to the passage you will find the same number and five words. Choose the word that makes the best sense in the blank.

On your answer sheet, find the same number as the blank. Fill in the circle that has the same letter as the word you have chosen. Mark only one answer for each question. If you want to change an answer, be sure to erase your first mark completely. Then mark the answer you want.

Find Samples 1 and 2 on your answer sheet. Read Sample 1 below and see how the right answer has been marked on your answer sheet. Then read Sample 2 and mark the answer on your answer sheet."

I will wait while you work on these two samples.

Do *not* read the samples aloud. Allow students time to do both sample questions themselves. Then say:

"Look at the answer for Sample 1. The circle for letter *d* is marked because the word *weather* makes the best sense in the blank.

In Sample 2 you should have marked the circle for letter *a* because the word *rain* is the word that makes the best sense in the blank.

As you can see, you may not be sure of the answer to a question until you have read the sentences that come after the blank. So be sure to read enough to answer the questions.

You are not expected to read at the same speed as other people or to answer the same number of questions. As you work on this test, you will find that the passages become harder to read. Do your best to read as many passages as you can and to answer as many questions as you can.

You will be given as much time as you need. Be sure to spend your time on what you can read rather than on what is too hard for you. It is more important to get the answers right for the passages you can read than it is to try to read all the passages on this test.

Remember, mark all your answers on the separate answer sheet, and mark only one answer for each question."

Do you have any questions? Do you know how to mark the answer sheet?

If students do not understand the sample questions or the marking of the answer sheet, the appropriate directions may be explained until everyone knows what to do. After all questions have been answered, say:

When you are finished working on the test, close your booklet and place it on top of your answer sheet. Do you have any questions? Now, turn the page and begin work.

If the test is administered in a regular classroom situation, students should remain quietly at their desks and be allowed to work on other assignments when they finish their tests. The examiner may collect the test materials either as students complete the test or when most of the students have finished.

If the test is administered in a large group situation, schools may prefer to allow students to hand in their test materials as they finish and then leave the room. If so, care should be taken that students leave the room as quietly as possible so as not to disturb the students who are still working on the test.

Scoring the Test

The student's score on the Reading Test for New York State Elementary Schools is the number of questions that the student answered correctly.

The list of correct answers is contained in Appendix I.

The NCS answer sheet may be either hand or machine scored. A scoring overlay is provided for hand scoring. On the overlay, boxes have been printed to indicate correct answers; when the overlay is aligned properly with the answer sheet, each correct answer will be

visible in a box. To handscore the test, use the following procedures:

1. Place the scoring overlay on the answer sheet. Be sure the overlay is correctly aligned so that the boxes correspond to the correct answers.
2. To obtain the total raw score, count the number of correct answers. If a student has marked two or more answers, omit the item in counting the number correct. Place the total raw score in the box provided on the answer sheet.
3. Convert the raw scores into DRP units through the use of the conversion tables provided on pages 23–25.

INTERPRETING TEST RESULTS

The State Reference Point and Remediation Requirements

The Regulations of the Commissioner of Education require schools to provide appropriate remedial instruction to students obtaining scores below a certain point on any of the Pupil Evaluation Program tests. Section 100.1 (f) of the Regulations states:

> Pupils who score below the designated Statewide Reference Point in the third-grade Pupil Evaluation Test shall be provided appropriate remedial instruction designed to enable them to score above the Statewide Reference Point on the sixth-grade Pupil Evaluation Program Test. Pupils who score below the designated Statewide Reference Point on the sixth-grade Pupil Evaluation Program Test shall be provided appropriate remedial instruction designed to enable them to score above the Statewide Reference Point on the preliminary competency tests and/or pass the Regents Competency Test in mathematics. The parent or guardian of a pupil who is to be provided remedial instruction shall be notified in writing, by the principal, of the Pupil Evaluation Program Test results and the plan for remedial instruction, which shall begin no later than the commencement of the school semester after which the test is given.

Accordingly, a Statewide or State Reference Point (SRP) has been established for the Reading Test for New York State Elementary

Schools used in the Pupil Evaluation Program. On Form E, raw scores of 51 or more are above the State Reference Point.

This SRP is considerably different from the reference point associated with the sixth-grade PEP reading test used prior to 1979, which was the 23rd percentile in the 1966 baseline norms. The new test is clearly criterion referenced, while the earlier tests were norm referenced.

The State Reference Point is a standard expressed in terms of readability. This point represents a 70 percent likelihood of success on prose materials typically found in textbooks used in grade six. Thus, schools are required to provide students whose scores are below the SRP with programs designed especially to meet their needs.

In accord with the mandate to identify students requiring remediation, placement groups have been constituted in relation to the

TABLE 1

PLACEMENT DECISIONS BASED ON THE READING TEST FOR NEW YORK STATE ELEMENTARY SCHOOLS GRADE 6 — FORM E

Number Correct	Probability of Comprehending Textbooks Used in Sixth Grade	Comment	Recommendation
56 or above	.76 and above	Well above the State Reference Point	Assign reading and writing tasks which challenge the student's ability
51 to 55	.70 to .75	Above the State Reference Point	Increase demands of reading and writing tasks in the regular instructional program
48 to 50	.65 to .69	Just below the State Reference Point	Provide remediation in the regular instructional program
40 to 47	.50 to .64	Far below the State Reference Point	Provide remediation beyond that included in the regular instructional program
39 or below	.49 and below	Very far below the State Reference Point	Provide intense individualized remediation

State Reference Point. Table 1 on page 9 lists the groups according to raw score (total number right) on the Reading Test for New York State Elementary Schools. It also indicates the recommended action that should be taken with respect to each group.

Notifying Students and Parents About Test Results and Placement Decisions

If students score below the State Reference Point, then students and parents must be informed of the student's raw score, its implications and the school's placement decisions. Students and their parents must also be informed of the content and scope of the reading programs that students are to receive. They should be advised about other reading activities that students should undertake on their own to improve reading skills. As a minimum, information similar to that which appears in letter A in Appendix II must be given both to the parents of students with raw scores of 48 to 50 and to the students themselves. For those students whose raw scores are 47 and below, information similar to that which appears in letter B in Appendix II should be provided.

Required Records

Schools are required to maintain individual records in all categories for students requiring special assistance, that is, those students scoring below the State Reference Point. These records should indicate the assistance provided over a specified time and should document progress in the ability to read with comprehension. It is prudent to keep on file copies of all correspondence sent to each student and his or her parents.

Recommendations: Students Scoring Above the State Reference Point

For students who achieve raw scores of 51 and above on the Reading Test for New York State Elementary Schools, no special reading instruction is required. However, in keeping with the goal of upgrading the reading and writing abilities of all students, these students should be given additional reading and writing tasks within the regular instructional program.

Recommendations: Students Scoring Below the State Reference Point

For purposes of planning appropriate instructional programs, students scoring below the State Reference Point may be viewed as constituting three groups:

1. *Students with raw scores of 48 to 50 on Form E.* Students in this group, if given consistent and sequential reading instruction within their regular subjects, will probably achieve the level of competency required for high school graduation.
2. *Students with raw scores of 40 to 47 on Form E.* Students in this group will need the type of assistance which goes beyond that given in regular subjects. Corrective instruction should be provided through well-organized supplementary programs.
3. *Students whose scores fall at or below 39 on Form E.* Students in this group exhibit serious reading deficiencies and will need intensive assistance typical of clinical remediaton on a very small group or individualized basis.

Using the Test for Instructional Purposes

A general discussion of DRP methodology is given below to aid the reader in using the test results for managing reading instruction, grouping students of comparable ability, deciding if a particular student is capable of reading text at the level required of students in sixth grade, and selecting text or other materials to match a particular student's instructional needs.

DRP tests measure a student's ability to process and understand nonfiction English prose passages written at different levels of difficulty or readability. The tests identify the hardest prose that students can read with comprehension.

The test items are designed so that the text of the passage must be read and understood in order for the student to respond correctly. All of the response options "fit" the blank space; each one makes a grammatically correct and logically plausible sentence if the sentence is considered in isolation. However, only one response "fits" or is plausible when the surrounding context of the passage is considered. Thus, the selection of the correct response depends on understanding the prose in the surrounding text. The student should be able to respond correctly to all (or nearly all) items in the test, up to the point where he or she cannot understand a passage sufficiently well to decide which word is the correct choice.

The deleted words and the other response options are words in common use and are of uniform difficulty throughout the test. Thus the test items become more difficult only with respect to the difficulty of the text in the passages. The response options are maintained at an easy, constant level so that students will not have difficulty recognizing the response options.

All of the information that is needed to select the correct response is provided in the text of the passages; there is no need to supply information concerning each passage's content from memory or past experience. This means that the DRP should be culturally fair. Knowing a great deal or nothing about the content of any passage should not bias a student's ability to respond. Only the ability to process sentences for meaning is required for success on the DRP.

The test is not norm-referenced, and no comparisons are made among students. The meaning of a student's score does not depend on comparisons with the performance of students in a norming sample. Instead, the tests measure a student's reading ability on an "absolute" scale. Just as a person's height and weight can be measured accurately without reference to how tall or heavy any other person is, so can reading ability be measured by determining on the prose difficulty scale the hardest text that can be read with comprehension.

Prose Difficulty or Readability

Prose varies in difficulty: it ranges from being very easy to understand at the one extreme, to being very hard at the other. This variability in prose is called readability.

While many factors determine the readability of prose, two or three factors have been shown to account for major differences in prose difficulty. The greater the proportion of common, frequently used words in a sample text, the easier the text is to read. The higher the frequency of simple, short sentences, the easier the prose. Therefore, the readability of prose can be measured by counting the proportion of common words and by calculating average sentence length.

A variety of readability formulas specify how to combine various factors or characteristics of prose. Some formulas (e.g., Spache and Dale—Chall) interpret readability in terms of the grade level at which material is most likely to be used. Although this practice produces simple readability scores, it very often leads to misunderstandings. In order to avoid these sources of confusion and to use

the most technically valid measures available, a new measure of readability is being introduced through the DRP tests. Readability is measured in Degrees of Reading Power (DRP) units.

The Measurement of Readability

DRP units form a scale of prose difficulty or readability. The scale ranges from 15 to 100 units. A score of 15 DRP units describes the easiest prose that anyone is likely ever to encounter. It is easier than text normally found in the English language. At the other extreme of the scale, a score of 100 DRP units represents the most difficult prose imaginable. In practice, commonly encountered English text runs from about 30 DRP units at the easy end to about 85 DRP units at the difficult end.

On the Reading Test for New York State Elementary Schools, Grade 6, Form E, the readability levels of the passages range from 38 to 81 DRP units. Table 2 is designed to help the reader understand the scale of DRP units by presenting the readability of various passages in DRP units.

In addition, a display of the range of readability of some well-known reading materials is presented in Figure 1. Figure 1 shows the average readability in DRP units of different types of articles sampled from child, teen, and adult magazines and from professional journals. Articles appearing in children's magazines contain easier prose than articles appearing in teen magazines. Teen magazine articles, in turn, contain easier prose than articles in adult general interest magazines. Professional journal articles contain the most difficult prose of all. In interpreting the readability means for the various article categories shown within the brackets in Figure 1, any difference of four or more DRP units may be considered a "real" or important difference.

Reading Ability and Prose Difficulty

DRP methodology enables predictions to be made concerning the difficulty of prose that a student can read with comprehension. The passages on the test are arranged in order of difficulty from easiest to hardest, and the difficulty of the test items is linked to the readability of the passages. Therefore, the greater the number of test items the student answers correctly, the harder the text the student can read with comprehension.

TABLE 2
The Readability of Prose Samples in DRP Units[1]

	Readability in DRP Units
A bird's wings are well-shaped for flight. The wing is curved. It cuts the air. This helps lift the bird. The feathers are light. But they are strong. They help make birds the best fliers. A bird can move them in many directions. Birds move their wings forward and down. Then they move them up and back. This is how they fly. The tail feathers serve as a brake. They also aid in steering.	38
A bark canoe proved to be better than a log canoe for several reasons. It hauled heavier loads. It glided more smoothly. A log canoe was hard to repair. But cracked, torn, or worn bark could be mended quickly. A bark canoe weighed less than a log canoe. It was so light a lone man could carry it. A log canoe could be no longer than a tree. But a bark canoe was different. It could be almost any length.	43
The part of a beach between high and low tide is called the middle beach. It is home to many plants and animals. But life on this middle beach is hard. There is no protection against the wash of the oncoming waves. Some animals survive by digging holes in the sand. They can stay in their homes under ground. The undertow will not pull them out to sea. They are safe.	47
Most creatures take great care to protect their eggs. The walking stick does not. It just drops its eggs, scattering them loosely on the ground. Dozens and dozens drop at a time. As the eggs fall onto dry leaves, they sound like raindrops falling. Many of the eggs do not hatch. But enough do so that the walking sticks will not die out. They have existed on earth since before the era of the dinosaurs.	51
Between 1935 and 1939, some 350,000 migrants crossed the border from Arizona to California. These migrants came mainly from Oklahoma and Arkansas. For this reason they were called Okies or Arkies. They came from areas badly parched by drought. Lack of rain, combined with strong winds, had meant the loss of countless tons of topsoil. The area was aptly called the "Dust Bowl."	55

[1]The readability calculations are based on longer samples.

TABLE 2 (Cont.)
The Readability of Prose Samples in DRP Units[1]

	Readability in DRP Units
The most ancient hand tool yet identified is the stone hand ax. Such instruments are more than 250,000 years old. Hand axes were made either from flint or from a fine-grained rock that would provide a sharp edge. Using a second stone, flakes were hammered off around the edges. The resulting tool had a jagged cutting edge down each side. It had a crude point and a smooth, round butt. During the thousands of years in which this tool was used, it underwent relatively slight changes in form	60
Wall paintings are especially vulnerable to atmospheric change. Archaeologists know this. Hence they try to discover, before opening a tomb, whether they will find murals. Special tools have been designed for this purpose. One of the most useful is a kind of camera that can be dropped into the ground before the digging starts. If the camera indicates the presence of wall art, scientists can prepare to take steps to preserve the painting as soon as it is reached.	64
The ancient Romans' concept of justice and their view of individual rights continue to be guiding forces in contemporary western society. Their code, developed by the emperor Justinian, provided, for example, that no person was punishable for his thoughts, that fathers could not be compelled to bear witness against their sons, nor sons against their fathers, and that the burden of proof rested with the accuser, not with the party denying the accusation.	72
Thomas Jefferson's belief in man's ability to govern himself is well known, but his views concerning the conditions under which this ability can be successfully realized are less familiar. Jefferson believed that a natural moral sense—which he assumed to be a universal characteristic of mankind—was the initial and basic prerequisite for self-government. The effective operation of this moral sense was then influenced, he believed, by environmental factors, in particular by economic circumstances.	81

[1]The readability calculations are based on longer samples.

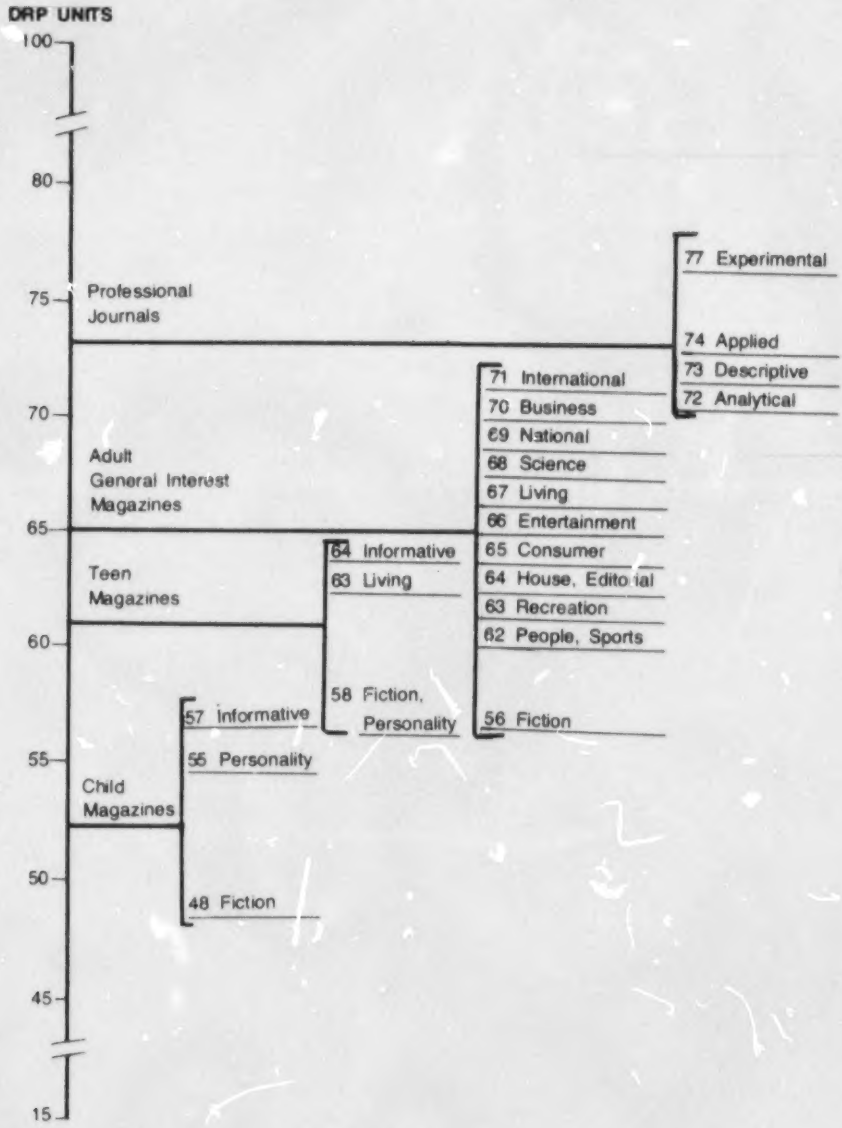

FIGURE 1.
The Average Readability of Sections of Professional Journals and High Circulation Adult, Teen and Child Magazines in DRP Units.

With DRP methodology, it is possible to convert the total number of items answered correctly (i.e., the total raw score) to a readability level, which is a prediction of the readability of prose that a student can read with comprehension. It is also possible to convert the total raw score into several different readability levels, or predictions of the readability of prose that a student can read with various likelihoods of success. The likelihood of success, or the degree of comprehension with which a student can read prose of a specified readability level, is provided for three specific and educationally important ability levels: independent reading level, instructional reading level, and frustration level. In order to provide the most appropriate instruction for a student, teachers must know which materials are useful at the student's independent level and instructional level, and which materials can be used under hardly any conditions because they are too difficult or frustrating.

The common way of defining these three levels has been in terms of the proportion of items that the student answers correctly on a test of reading comprehension. Materials at the independent level are commonly defined as those on which the student can answer correctly at least 90% of the items. Materials at the instructional level are commonly defined as those on which the student can answer correctly at least 75% of the items. Materials on which the student can answer only 50% of the items are commonly considered to be at the frustration level. An alternative way of stating these definitions is that the independent, instructional and frustration levels are, respectively, those prose readability levels at which test items would be answered with a .90, .75, and .50 likelihood of success. Through the use of DRP methodology, the test score can identify these three levels for each student. It should be kept in mind that the ability to read material at the independent level rather than the instructional level represents on the Dale-Chall readability scale a difference of from 3 to 4 grade levels in the middle of the scale.

Converting Raw Scores to Reading Ability Levels

The DRP raw score does not by itself provide information concerning what materials the student can read with a given likelihood of success. To give meaning to raw scores they must be converted to DRP units. Table 3 is used for this purpose. In this table, raw

scores are translated into predictions concerning students' ability to read prose materials that have different readability values.

The first column at the left of Table 3 gives the raw score, that is, the total number of questions answered correctly on the Reading Test for New York State Elementary Schools. The entries in the remaining columns are expressed in DRP units. For any given raw score, the column labeled "Independent Level (p=.90)" gives the readability of materials in DRP units for which a student with that raw score has a .90 likelihood of success. Since such a student would also be expected to answer correctly at least 90% of questions based on any easier materials, the tabled entry defines the most difficult prose suitable for the independent level. Any easier prose would also fall within the student's independent level.

The set of columns labeled "Instructional Level (p=.75)" gives the readability of materials in DRP units for which students have about a .75 likelihood of success. Because the instructional level is more properly thought of as a range of difficulty rather than a single point, the table shows the range as one standard error (4 DRP units) on either side of the middle value. Thus a lower and upper boundary of the instructional level are defined for each raw score.

The extreme right-hand column entry labeled "Frustration Level (p=.50)" identifies the readability of materials on which students have a 50-50 chance of responding correctly. This level of performance traditionally has been defined as the frustration level for the student. Such materials would be too difficult for most students to use profitably. However, strong readers may want to attack these materials, and they should not be prohibited from doing so. The entry shown in the extreme right-hand column defines the point at which the frustration level begins.

As an example, Table 3 indicates that a student who achieves a raw score of 58 can read and comprehend at the independent level prose materials with a readability of 48 DRP units or below. The middle value of the instructional level for the same student falls at 59 DRP units. The upper and lower bounds for the instructional range are 64 and 54 DRP units, respectively. Therefore, the instructional level of materials for this student can be as easy as 54 DRP units or as difficult as 64 DRP units. This student may be expected to comprehend material within this range of readability, with instructional support. The final column entry for this same student indicates that materials with DRP unit values of 70 and above will be at the frustration level.

Note that in some columns there are no entries for the very high and very low raw scores. At the high end of the table, a student with a raw score of 76 can read independently materials with a DRP unit value of 84 which is very difficult text. At the easy end of the table, cells are left empty when there are no materials easy enough to fill them. For example, a student with a raw score of 10 is a poor reader. The tabled values show that there are no materials easy enough for this student to use at the independent level. Materials with a DRP unit value of 19 to 24 are within such a student's instructional level, but the range is bounded only at one side: there are no materials easy enough to define a lower bound or "easy" end for this student's instructional level. Fifteen DRP units defines the lower end of the DRP scale and represents the easiest possible prose that might be found in English.

There is no row of DRP unit values for a perfect raw score of 77. This occurs because the statistical procedures used for making predictions do not apply to a student who answers all the items on the test correctly. In effect, since such a student has not been given any questions that he or she could not answer, the limit of this student's ability is unknown. Hence, specific predictions cannot be made concerning which materials are suitable for which purposes for students scoring 77. It is safe to assume that no materials commonly used for the sixth grade will pose significant reading problems for such a student.

To help teachers use the results of the Reading Test for New York State Elementary Schools to match student ability to the difficulty of instructional materials, the Department has prepared a publication entitled "Degrees of Reading Power Readability Report." This publication provides DRP unit readabilities for many of the textbooks currently in use in schools in grades 4–12. Listings of textbooks are arranged by content area and grouped within content area according to readability level. By using these readability listings together with a measure of student ability in DRP units, teachers will be able to prescribe the appropriate match between student ability and the difficulty of instructional materials.

Work on computing the readabilities of instructional materials commonly used in New York State schools is continuing, and the results will be published in supplements to this publication. Schools that are interested in obtaining analyses of instructional materials not included in this publication may request them by using the form provided in Appendix IV, or a facsimile of it. The requests will be

catalogued and analyses performed first on those materials with the widest potential for use.

As described in the Readability Report, the DRP unit readabilities are computed following sophisticated and rigorously established procedures which are considered necessary in matching student ability with the difficulty of materials, in order to reduce to a minimum any error that might be attributable to the unreliability of the readability index. Prior to the DRP tests and methodology, less-rigorous procedures might have been acceptable because no precise estimate of student abilities was available to use in matching ability and difficulty.

Until the readabilities of most of the instructional materials used by schools have been computed in DRP units, interim procedures can be used to provide some degree of match between student ability and the readability of instructional materials. Using the passages in Table 2 as a guideline, it is possible to make judgments about the reading difficulty of materials not yet included in the Readability Report. This can be done by dividing the book into tenths and comparing the text of each tenth of the book to the passages contained in Table 2. When estimated readability values have been assigned to each tenth of the book, it is possible to determine the range of difficulty of materials as well as the rate of increase in difficulty from the beginning to the end of the book.

Students with similar DRP scores can then be assigned to appropriate groups within classrooms or, in the case of team teaching, to particular "classes," and appropriate materials selected for them. Since the instructional levels listed in Table 3 overlap for several successive raw scores on the test, it is easier for schools to group students and for teachers to select students for small group instruction.

Using the Test as a Diagnostic Instrument

The Reading Test for New York State Elementary Schools is especially designed to aid in the management of instruction. It provides the diagnostic information needed to match student ability to instructional materials, a feature which is considered by many psycholinguists to be the most important diagnostic capability of a test. Matching allows a teacher to control the amount of challenge in various reading situations. A teacher may desire challenge in materials supplemented by instruction but may prefer materials that

students can comprehend without help for independent reading. Students who might become bored or frustrated with the usual instructional materials can be identified immediately with the test, and their reading assignments changed accordingly. Both the amount of learning gained from instructional materials and the motivation to read will be greatly affected by proper matching of student ability to text difficulty.

Reading specialists can also use the test diagnostically to identify problems individual students have with the comprehension of prose. A student's test responses can be studied and the student questioned concerning why a particular incorrect answer was selected. This probing procedure will be most useful if the materials are in range for a student; that is, if the student can correctly answer about 70–75% of the questions.

Because of the unique characteristics of the test, it is also possible for teachers to derive diagnostic information by examining the pattern of a student's answers. For example, if a student receives a score of 5 or less on a passage and scores 6 or 7 on subsequent passages, then carelessness can be assumed on the easier passage. If the student misses most of the questions on the beginning passages and then improves, guessing is a possible reason.

As noted above, the DRP can provide diagnostic information if students' responses to test questions are carefully probed. Basal readers and other texts are also sources of diagnostic information when suitable probes are made with materials that are at the student's instructional level. Since the DRP forecasts which books are at the instuctional level, it can be used to identify materials that should yield good diagnostic information when key questions are asked concerning what students understand in the text.

Rather than formulating their own questions, many teachers obtain diagnostic information by using the multiple choice (or other) questions that appear in basal readers and other instructional materials. The best information from such exercises will be obtained if the questions that are asked are based on prose known to be at the instructional level. Since the DRP identifies such prose, its use in selecting materials for students increases the likelihood that teachers will obtain useful information from the diagnostic exercises.

Diagnosing students' reading abilities with questions that appear in instructional materials rather than with conventional diagnostic tests has certain advantages. First, when teachers use classroom materials, they can control the readability of the materials about which

a student is being questioned. This is not true of conventional tests, where some prose may be too easy or too difficult for the student. Furthermore, when diagnosis is based on classroom materials, the teacher can make a large number of observations at the appropriate ability level. This cannot be done when conventional tests are used, since they are designed to assess a wide range of abilities and hence provide very few items which can be used diagnostically at each ability level. Being able to make more observations, and also being able to space them over time, should produce a more reliable diagnosis. Finally, when actual instructional materials are used there are no issues concerning whether diagnostic test results apply to the teaching situation.

Unfortunately, the difficulty of diagnostic questions—whether asked in the classroom or on a conventional reading test—is not necessarily linked to the difficulty of the prose on which the questions are based. A question that is asked about text may be much harder or more abstract than the text itself. Therefore, caution is required in interpreting the results of all diagnostic questions, whether they be exercises printed in instructional materials, test items, or probes.

Organizing Reading Programs

In organizing programs both to refine the skills of able readers and to provide remediation designed to correct reading deficiencies, the following concepts should be given consideration:

- The development of reading ability and the remediation of reading deficiencies is a responsibility of all teachers. Reading instruction should be included in the teaching of all subjects.
- The further students are below the State Reference Point, the more intense the instruction needed for them to learn to read with comprehension.
- Students make the most progress by reading prose that falls within their instructional level.
- Comprehension at the level at which students are able to grasp, reorganize, and restate the message of written materials is a fundamental prerequisite for more advanced levels of comprehension and for the development of understanding at higher interpretive or critical levels. It is therefore good practice to implement instructional models that require active processing of materials at students' instructional levels.

TABLE 3

Predictions From DRP Raw Scores to the Readability of Prose (in DRP units) That Can Be Read at Different Comprehension Levels
READING TEST FOR NEW YORK STATE ELEMENTARY SCHOOLS

Grade 6—Form E

Total Items Right	Independent Level (p=.90)	Instructional Level (p=.75) Lower Bound	Middle Value	Upper Bound	Frustration Level (p=.50)
76	84	90	95		
75	76	83	88	92	99
74	72	79	83	88	94
73	69	75	80	85	91
72	66	73	77	82	88
71	64	70	75	80	86
70	62	69	73	78	84
69	60	67	72	76	83
68	59	65	70	75	81
67	57	64	69	73	80
66	56	63	67	72	78
65	55	61	66	71	77
64	54	60	65	69	76
63	53	59	64	68	75
62	52	58	63	67	74
61	51	57	62	66	73
60	50	56	61	65	72
59	49	55	60	64	71
58	48	54	59	64	70
57	47	53	58	63	69
56	46	52	57	62	68
55	45	52	56	61	67
54	44	51	55	60	67
53	44	50	55	59	66
52	43	49	54	58	65
51	42	48	53	58	64
50	41	48	52	57	63
49	40	47	52	56	63
48	40	46	51	55	62
47	39	45	50	55	61

TABLE 3, cont'd

Total Items Right	Independent Level (p=.90)	Instructional Level (p=.75) Lower Bound	Middle Value	Upper Bound	Frustration Level (p=.50)
46	38	45	49	54	60
45	37	44	49	53	60
44	37	43	48	52	59
43	36	42	47	52	58
42	35	42	46	51	58
41	35	41	46	50	57
40	34	40	45	50	56
39	33	40	44	49	55
38	32	39	44	48	55
37	32	38	43	47	54
36	31	38	42	47	53
35	30	37	41	46	53
34	30	36	41	45	52
33	29	35	40	45	51
32	28	35	39	44	50
31	27	34	39	43	50
30	27	33	38	42	49
29	26	32	37	42	48
28	25	32	36	41	47
27	24	31	36	40	47
26	24	30	35	39	46
25	23	29	34	39	45
24	22	29	33	38	44
23	21	28	32	37	44
22	20	27	32	36	43
21	20	26	31	35	42
20	19	25	30	34	41
19	18	24	29	34	40
18	17	23	28	33	39
17	16	22	27	32	38

TABLE 3, cont'd

Total Items Right	Independent Level (p=.90)	Instructional Level (p=.75) Lower Bound	Middle Value	Upper Bound	Frustration Level (p=.50)
16	15	22	26	31	37
15		20	25	30	36
14		19	24	29	35
13		18	23	28	34
12		17	22	26	33
11		16	21	25	32
10			19	24	30
9			18	22	29
8			16	21	27
7				19	26
6				17	24
5				15	22
4					19
3					16
2 and below		*** All Values Below 15 ***			

- Instructional programs should require that students write about what they have read. This facilitates the development of both reading and writing skills.
- Students are motivated to read when given reading materials both appropriate for their level of reading ability and in accord with their interests. A wide variety of reading materials should be available in schools and classrooms to increase the likelihood that materials appropriate in difficulty will be used with every student.

Programs based on these concepts require a significant commitment of school resources, instructional time, and staff effort. It is only through such commitment on an organized and continuing basis that the reading skills of all students can be developed.

Using the Test as a Basis for Terminating Remediation

Form E may be used to assess the reading ability of students who were identified as needing remediation at the end of grade 6. The test may be administered at the middle or end of grade 7 to help determine whether a remedial program should be continued. It is recommended that pupils obtain a raw score of at least 35 on Form E at the middle or end of Grade 7 before the school considers terminating their remedial programs.

TECHNICAL CHARACTERISTICS

The Validity and Reliability of the Test

The DRP is scored to provide a sample-free (i.e., non-normative) estimate of students' reading ability. Conceptually, the passages in the DRP may be thought of as a series of hurdles of increasing difficulty. According to the theory of the test, the student correctly responds to items, or clears the hurdles, until the point is reached at which the student can no longer understand the prose in the passage sufficiently well to decide which response option is correct. At this point, the limit of the student's reading ability has been reached.

A scoring model that corresponds to the concept of the DRP is the Rasch model, a latent trait model with two parameters: the ability of the person taking the test and the difficulty of the test items. According to the Rasch model, the probability of a correct response to a test item is a logistic function of the difference between the difficulty of that item and the ability of the person responding. The model provides the mathematical basis for calibrating both items and people in terms of difficulty and ability, i.e., for obtaining unbiased estimates of the model parameters and for determining the probability of success on items for people of different abilities. The logic of the Rasch model is consistent with the design and structure of the DRP. Students may be expected to perform successfully on items or passages well within their ability and fail items or passages having difficulties well beyond their ability.

DRP tests fit the Rasch model according to the relevant statistical tests for both the unconditional maximum likelihood and the conditional maximum likelihood approaches. The conditional likelihood approach is used with the DRP because it allows item bias by ability group, by sex, and by ethnicity to be checked for each raw score and for all abilities combined, in ongoing calibration studies. If an item turns out to be biased, the item and passage are revised and recalibrated.

A number of studies have established the reliability and validity of the DRP. Some studies are ongoing, such as the tests for interactions between score groups, sex, and ethnicity. These are executed every time new passages are entered into the passage bank. Other studies are done periodically.

In the following section, a selected set of completed studies is summarized. These studies and experiments involved approximately 5,000 fourth- and sixth-grade students. For these studies and experiments 24 DRP passages were prepared, each containing 7 items. There were 2 passages at each of 12 levels of difficulty ranging from extremely easy prose to quite difficult prose. The 2 passages at each level of difficulty were then randomly assigned to one of two parallel test forms. The students, drawn from schools where the minority population varied from 20 to 80 percent, were assigned at random to participate in one of several studies which are briefly reported on pages 28–32. These studies are reported in greater detail elsewhere. It should be borne in mind that the results reported are based on elementary school populations; the reliabilities of DRP tests centered for high school populations are as high or higher.

Traditional Reliability and Validity Studies. Students took the two randomly constituted parallel forms of the DRP with a two-week interval between the administration of the forms. The order of administration was counterbalanced. Product-moment correlations showed that the DRP is reliable according to the constructs of True Test Score theory. For the randomly constituted parallel forms, the reliability coefficients were $r = .86$ in grade 4 and $r = .91$ in grade 6. The DRP is also homogeneous, with $KR_{20} = .95$ for each form. Thus, by conventional standards, the DRP is reliable.

The DRP also meets conventional standards for validity. Both convergent and discriminant evidence were found for the construct validity of the test. In different studies students took either the DRP and a norm-referenced reading achievement test or they took the DRP, two kinds of cloze tests, and a general aptitude test. Results of these studies showed that the DRP correlates more with the language subtest of an aptitude measure ($r = .80$) than it does with the non-language subtest ($r = .70$). The DRP, as expected, has moderately high correlations with an "any-word" cloze test ($r = .76$), with a "hard" cloze test ($r = .78$), and with a conventional norm-referenced reading test ($r = .84$). In accord with expectations, the DRP is less strongly related to a measure of vocabulary knowledge of words presented in partial isolation ($r = .56$) than it is to traditional multiple-choice comprehension questions ($r = .80$).

Nontraditional Validity and Reliability Studies. Several studies were designed to evaluate whether the DRP is valid and reliable in accord with the meanings of these terms in the natural sciences.

A. *Construct Validity.* DRP items are constructed so that all response options make semantic sense when the items are taken out of the text in which they are ordinarily embedded. Isolated DRP items do not contain sufficient semantic information to make one item more or less difficult than another. DRP item difficulties depend entirely on the readability of the prose in which they are embedded. Studies have determined that the average item difficulty per DRP passage is accounted for by the difficulty of the surrounding prose.

In one study the difficulty of DRP items changed when the difficulty or readability of the surrounding text was altered. The fact that the difficulties of identical items varied as a function of the difficulty of the prose in which they were embedded demonstrates that DRP item difficulties measure the readability of prose.

A multiple correlational study determined whether the surface linguistic variables known to account for readability would predict DRP

passage difficulties. The mean and median Rasch item difficulties were calculated for each passage. These average difficulties became the criterion scores in two multiple correlations. The predictor variables were the proportion of hard words, the average word length, and the average sentence length per passage.

The multiple R's between the predictor variables and mean and median passage difficulty were, respectively, R=.98 and R=.96 for one form, and R=.95 and R=.96 for another form. Thus, essentially all the variance in average passage difficulty was accounted for by the surface variables. As well as can be measured, the mean and median Rasch item difficulties have no meaning other than indexing the difficulty of prose. Since Bormuth's (1969) mean cloze readability formula (see the Readability Report) employs the same predictor variables, the simple correlation between forecasts of DRP passage difficulty based on the Bormuth formula and their actual values should also be high. As expected, DRP passage difficulties for both forms correlated $r = -.98$ with Bormuth's measure of readability.[2]

On the DRP, the greater the number of items a student answers correctly, the higher the ability of the student. Since it has been shown that DRP item difficulties depend entirely on the difficulty of the prose in which they are embedded, it follows that we have shown that the higher a person's ability as demonstrated on the DRP, the more difficult the prose the person can read with comprehension.

The near-perfect correlation between the average item difficulties of the DRP passages and the Bormuth readability values for the passages allows relatively precise cutoff scores to be calculated when expectations or goals are formulated in terms of readability requirements. Once a readability goal is formulated, the ability required to comprehend at that readability level can be predicted by using the regression equation for predicting ability from readability.

Since the probability of success on an item is a function of the difference between item difficulty and ability, it is possible to calculate the probabilities that people of different abilities will be able to answer questions at the goal. When the user sets both the readabil-

---

[2] Note that the Bormuth mean cloze scores form a scale on which high scores indicate easy materials, while Rasch values increase with difficulty: hence the negative correlation between the Bormuth scores and mean and median passage difficulty. Since Bormuth's (1969) mean cloze formula is used to derive DRP units, as explained in the *Readability Report*, the magnitude of this correlation is important. Refer to Appendix III for derivation and validity of DRP units.

ity level and the desired probability of success, then it is possible to express the goal in terms of the total number right on a particular form.

The process may be reversed. It is possible to forecast the probability that a person with given DRP ability will be able to read text at a given level of readability. The forecasts at probabilities of .90, .75, and .50 are displayed in the conversion tables for each form of the DRP (e.g., see Table 1).

B. *Predictive Validity.* Manipulations with the Rasch model enable forecasts to be made concerning the difficulty of prose material that can be read with various likelihoods of comprehension (success) by students of different ability levels. The validity of these forecasts is supported by the fact that DRP data fit the Rasch model and by the fact that average DRP passage difficulties measure readability. Even so, an experiment was undertaken to validate forecasts concerning how well students can read ordinary prose materials.

The DRP was administered to one group of students (the controls), and forecasts were made concerning the readability level of prose passages that students should be able to read with .90 and .75 likelihood of success. Hard cloze tests, made by deleting every fifth or next eligible word that was a noun, verb, adjective, or adverb, were used to evaluate whether prose at a given level of difficulty could be read with the predicted amount of success. Because of task differences, it was necessary to translate the 90 and 75 percent predictions on multiple-choice tests into proportions correct on hard cloze tests. The requisite data were obtained in a separate study using the same population of students. It was found that 27.4 and 16.7 percent correct on hard cloze was the expected value for 90 and 75 percent correct on multiple choice. When the actual proportion correct on the hard cloze tests was compared with expectations, the expectations were confirmed.

Another group of students (the experimental condition) also took the DRP. They were then asked to read the intact passages on which the hard cloze tests were based, immediately before they took each of the hard cloze tests. A criterion level of 31.2 percent correct on the hard cloze and 95 percent correct on multiple choice was established as an indicator of mastery. Except for the lowest ability group, all others were at the mastery level only on those passages that, according to DRP forecasts, they had at least a .75 likelihood of comprehending. Thus, it was shown that giving students intact text to read just before they took a hard cloze comprehension test led to

"mastery" level performance if, and only if, the readability of the test was at the instructional level (P=.75) relative to the student's ability as predicted by the DRP.[3]

Because there are many sources of error in evaluating predictions using "mastery" or other cutoff scores, other predictions were also evaluated.

The only difference between the experimental and control conditions was that the students in the experimental condition had the opportunity to read the intact passages before taking the cloze tests. Thus students in the experimental condition had an added opportunity (compared with the control subjects) to acquire or gain information from reading the passages, since the control subjects had the opportunity to acquire information only from the deleted-word versions (hard cloze tests) of the passages. Therefore, as might be expected, the experimental subjects, regardless of ability, always did better than the controls on the hard cloze tasks. However, as the passages became more difficult, the performance on the cloze tests of the experimental group converged with the performance of the control group. That is, the difference between the amount of information the experimental group was able to acquire from reading the intact passages and the information the control group was able to acquire from the cloze tests alone grew smaller as passage difficulty increased.

In accord with predictions, the higher students' ability on the DRP, the better their performance on the hard cloze tests in both the control and experimental conditions. DRP scores correctly forecast the performance of students of all abilities.

Because the probability of success should decrease rapidly as the discrepancy between prose difficulty (readability) and student ability increases, it was possible to demonstrate DRP predictive validity in two other ways. First, while cloze test performance generally declined as the passages grew more difficult, the cloze test performance of the high ability students decreased more slowly. Second, the higher the ability of the students, the slower the rate of convergence between the experimental and control conditions. That is, the DRP successfully predicted the difference between high- and low-scoring students in their ability to acquire and use information from increasingly difficult passages before taking the cloze test.

[3] These results suggest that when the aim of instruction is to enable individuals to master materials in terms of assimilating content, the conventional wisdom is correct: the materials used should fall within an individual's instructional level.

C. *Evidence that the DRP is Culture Fair.* Using the data collected from the information acquisition study described above, a set of demanding tests for bias were performed. Students who received the same DRP scores acquired the same amount of information from ordinary prose regardless of their sex, ethnic background, or socioeconomic level. The results of these analyses showed that DRP ability scores do not have to be qualified by knowledge of sex, race, or socioeconomic status in making predictions concerning student performance on hard cloze tests taken either alone or after intact prose passages have been read.

D. *Gain and Reliability: A Natural Science Approach.* Evidence was obtained that DRP tests can measure *individual* gain over a five-month interval even when broad-band test forms covering a wide range of readability are used. Two parallel forms of the DRP were administered with a five-month interval between tests. Students who took one form as the pretest were administered the other form for the posttest. All ability levels gained, and this gain was statistically significant for all ability levels in grade 6 and for all but the highest ability group in grade 4.

Reliability may be assessed by determining whether within-individual performance is the same over a time period when no change in ability is expected. When the parallel forms of the DRP were administered with a two-week interval between the tests, no differences, other than those due to guessing at the lowest ability level, were observed between scores. The DRP does make the same decision concerning ability with repeated measurements.

Given that the DRP makes the same decision concerning ability with repeated measurements, and that students differ in their abilities, it follows that the correlation between the pretest and the posttest should be high. As reported earlier, it was $r = .86$ in grade 4 and $r = .91$ in grade 6.

Unlike the test for reliability, the test for gain determined whether within-individual change was obtained over a period of time in which change was expected. Both tests are in accord with the constructs of gain and reliability as they are elaborated in the natural sciences.

Grade-Equivalent and Grade-Level Readability Scores

With few exceptions, grade-equivalent scores are usually provided for standardized achievement tests. However, for several reasons, it is not desirable to provide this type of score.

The first problem with grade-equivalent scores is that no meaningful "average" or typical score exists for any given grade. In a given grade, when the number of students getting each score is plotted, the curve is almost never symmetrical about some "average" raw test score. for grade-equivalent scores to be meaningful, a typical score for a grade is needed.

The second problem with grade-equivalent scores is that they lead to misunderstandings. There is a tendency to confuse the grade-equivalent score with a desired or ideal standard of performance, which it is not. There is also a tendency to make the judgment that students are doing satisfactory work if they are performing up to the "norm."

The third major problem with grade-equivalent scores is that they may impede good educational practices in reading instruction. In making reading assignments, it is important to know the difficulty of materials a given student can read with and without help, and to know which materials are beyond the ability of the student. The grade-equivalent score may be mistakenly assumed to designate the grade-level text that should be assigned for reading instruction. the grade-equivalent score explicitly does not have this meaning. Even though highly skilled reading teachers may, on the basis of experience, develop a way to predict what materials a student can read from the student's grade-equivalent score, such predictions are bound to be wrong on occasion, because scores provided on norm-referenced tests were simply not designed to say anything about which materials a student can read.

In addition, there are a variety of reasons why grade-level readability scores are not provided. First, grade-level readability scores would be readily confused with grade-equivalent reading scores. Second, grade-level readability scores are based on informal "normative" methods, and there is no longer a need to rely on such informal methods. Third, the most technically adequate and valid method to measure readability across the *entire* range of prose difficulty does not express readability in terms of grade levels.

APPENDIX I

Lists of Correct Answers for Form E

Form E

1 b	22 b	43 b	64 b
2 c	23 a	44 a	65 c
3 a	24 b	45 e	66 b
4 b	25 c	46 c	67 a
5 c	26 a	47 d	68 e
6 e	27 b	48 a	69 a
7 a	28 d	49 c	70 d
8 b	29 c	50 b	71 b
9 c	30 a	51 b	72 c
10 a	31 e	52 a	73 e
11 e	32 c	53 a	74 b
12 b	33 d	54 c	75 a
13 b	34 e	55 d	76 c
14 e	35 a	56 e	77 a
15 e	36 b	57 c	
16 a	37 d	58 d	
17 d	38 a	59 b	
18 e	39 d	60 a	
19 b	40 c	61 b	
20 c	41 e	62 e	
21 a	42 b	63 c	

APPENDIX II
Sample Letters to Parents

Letter A

The sample letter below contains the type of information that schools must provide the parent(s) or guardian(s) of students with scores below the State Reference Point (students with scores of 48 to 50). When appropriate, the letter should be written in the native language of the parent(s).

Dear (Mr. and Mrs. Smith):

Your (son John) took the Reading Test for New York State Elementary Schools on (date) and correctly answered (49) of the 77 questions on Form E of the test.

This test score indicates that (John) does not yet read well enough to be likely to pass the Regents competency test in reading which (he) will be taking in grade 11 or grade 12. (He) must pass that test in order to receive a high school diploma. It is very important, therefore, that (John) continue to improve (his) reading ability.

As we work with (John) on improving (his) reading, you can also help. One way (John) can improve (his) reading is to do a lot of it. You, therefore, can help by encouraging (him) to read at home. (He) should read whatever interests (him); the more (he) reads, the better (he) will be able to read.

If you want to know more about the Reading Test for New York State Elementary Schools or how to help (John) read better, please call ( ) at (telephone number).

Sincerely yours,

(School Administrator)

Letter B

The sample letter below contains the type of information that schools must provide the parent(s) or guardian(s) of students with scores far and very far below the State Reference Point (students with scores of 47 or below on Form E). When appropriate, the letter should be written in the native language of the parent(s).

Dear (Mr. and Mrs. Smith):

Your (daughter Mary) took the Reading Test for New York State Elementary Schools on (date) and correctly answered (46) of the 77 questions on Form E of the test.

(Mary's) test score indicates that (she) cannot read many of the books commonly used in elementary school courses. (Mary) may not be progressing at a rate that will allow her to pass the Regents competency test in reading, and (she) must take and pass this test in grade 11 or grade 12 in order to receive a high school diploma. It is, therefore, very clear that (Mary) must improve (her) reading ability.

In order to help (Mary) improve (her) reading ability, we are placing (her) in a special reading program called: _____. A description of this program is enclosed with this letter.

As parents, you can also help (Mary) improve (her) reading skills. You can encourage (her) to do a lot of reading at home. The more (she) reads, the more (her) reading will improve. (She) should read whatever interests (her) and should read as much as possible.

If you have any questions about (Mary's) test score or (her) special reading program, or about how you can help improve (her) reading, please call ( ) at (telephone number).

Sincerely yours,

(School Administrator)

Enc.

APPENDIX III
Derivation and Validity of DRP Units

Derivation and Validity of DRP Units

1. *The Derivation of DRP Units*

DRP units index the readability of prose. They are derived from the following mean cloze formula developed by Bormuth (1969):

For passage X, let the number of:

a) letters in passage = LET,
b) words in passage = W,
c) Dale Long List words in passage = DLL, and
d) sentences in passage = SEN, then

$$\text{Readability (R)} = .886593 - .083640\,(\text{LET/W}) + .161911\,(\text{DLL/W})^3 - .021401\,(\text{W/SEN}) + .000577\,(\text{W/SEN})^2 - .000005\,(\text{W/SEN})^3$$

Readability (R) in this formula is a cloze score, or the proportion of correct restorations that would be expected if a large sample of students, with widely varying abilities, were to be given cloze tests on passage X. Cloze tests are made by deleting words from text according to a mechanical plan and asking students to restore the missing words. The greater the proportion of correct restorations, averaging over all reading abilities, the easier the text.

Bormuth's method for developing the formula was to select at random 330, 110-word prose passages, develop cloze tests by deleting every fifth word, and then administer the tests (using all 5 forms of all passages) to matched samples of students in grades 4–12. The average proportion of exact restorations, ignoring misspellings, was calculated for each passage over all students (about 2,600). These average cloze scores became the criterion scores of passage difficulty for the readability formula.

Bormuth then found a more-or-less efficient set of linguistic variables which predicted the criterion scores via multiple regression techniques. As in other research *not* using cloze scores as the criterion, he found that the proportion of "easy" words (words on the Dale Long List), average word length, and average sentence length predicted the difficulty of the passages as known from the average cloze scores.

Finally, Bormuth tested the adequacy of his equation by: a) predicting the readability of a new set of 250- to 300-word passages; b) giving cloze tests made from the new passages to a sample of students; and c) determining the correspondence between the actual

mean cloze results and predictions. The correlation ($r = .92$) was acceptably high: the formula accurately predicts cloze scores for 250- to 300-word passages.

Bormuth's readability formula predicts cloze scores from a set of linguistic variables. Bormuth's R is a proportion: the larger the proportion, the easier the prose. On the DRP test of reading ability, though, the higher the total number right, the greater the student's ability. In order to make readability measures and reading ability scores run in a consistent direction (and to convert to percents), R must be transformed. Accordingly, DRP units are defined as follows: DRP units = (1-R) × 100. The higher the DRP units, the more difficult the prose.

2. *The Validity of DRP Units*

In the literature, the Bormuth formula is regarded as technically superior to other formulas for several reasons. First, use of the cloze technique to derive criterion scores is considered desirable because cloze provides a *prima facie* link between the difficulty of the prose and the difficulty of test items based on that prose. By comparison, the criterion scores employed in building many widely used formulas were based on responses to multiple-choice questions asked about prose passages.[4] A troublesome problem with measuring prose difficulty with this method is that there is no way to know whether the questions themselves are harder or easier, or more or less abstract in meaning, than the passage itself.[5]

In addition, the Bormuth formula: a) provides a continuous scale over the entire range of readability of first-grade through at least twelfth-grade instructional materials; b) has higher validity in original development ($r = .83$ for 110-word passages) than the best of other formulas; c) has a higher cross-validation coefficient ($r = .92$ for 250- to 300-word passages); d) has relatively low standard errors (S.E.) of measurement (S.E. = .06) in original development,

---

[4] Some formulas have used passage difficulties derived from existing formulas as their criterion. Since the formulas themselves are known to contain an appreciable amount of measurement error, they are of doubtful value as a criterion for building new formulas. For example, passage difficulty values from the McCall-Crabbs Standard Test Lessons in Reading have been used as criterion scores for readability formulas. These lessons were never intended for use as a criterion measure and were never rigorously standardized. Grade-level values based on these McCall-Crabbs passages may lack reliability and comparability (Stevens, 1980).

[5] DRP test items, unlike ordinary multiple-choice questions, are demonstrably linked to the difficulty of the prose.

S.E. = .04 in cross-validation); and e) has published cross-validation data.

No readability formula could be expected to forecast prose difficulty without error. Bormuth's R accounts for some 85% of the variance in observed cloze scores when 250- to 300-word passages are analyzed. Therefore, Bormuth's readability formula will not predict the difficulty of prose perfectly. However, cloze has been shown to be a good predictor of readability, and Bormuth's R, expressed in DRP units, does explain a great deal of the variation in prose difficulty.

Since all readability analyses in Bormuth's report are based on 250- to 300-word passages, the standard error of estimate is 4 DRP units. Therefore, a difference of 4 or more DRP units is statistically meaningful.

References

Bormuth, J. R. *Development of readability analyses.* Final report, Project Number 7-0052, Contract Number OEG-3-7-070052-0326, Office of Education, Bureau of Research, U. S. Department of Health, Education, and Welfare, March, 1969.

MacGinitie, W. H. and Tretiak, R. Sentence Depth Measures as Predictors of Reading Difficulty. *Reading Research Quarterly*, 1971, *6*, 364-377.

Stevens, K. C. Readability Formulae and McCall-Crabbs Standard Test Lessons in Reading. *The Reading Teacher*, 1980, *33*, 413-415.

APPENDIX IV
Request for Readability Analysis

REQUEST FORM FOR READABILITY ANALYSIS

An integral component of the DRP methodology is the matching of student ability to the difficulty of instructional materials. To aid in this matching, the Department is continuously analyzing commonly used textbooks for readability.

Books that are appropriate for analysis are those that are used in the classroom that consistently have passages of 250 words or more throughout the book. Mathematics books and spelling textbooks are generally not appropriate for readability analysis.

If there are books you would like analyzed that do not appear on the current readability report, please supply the information requested on the following page. A separate request is required for each book title; therefore, each book of a series must be requested separately. It is essential that the information supplied be complete and correct. The title should be written exactly as it is on the book. Do not omit any information. Persons requesting books for analysis may be asked to loan the books to the Department. The most frequently cited textbooks will appear in the readability analysis to be published each fall.

PLEASE PRINT CLEARLY

Book Title _____

Publisher(s) _____

Author(s) _____

Copyright Date: _____

Subject Area _____

Grade(s)in which used _____ Number of pages _____

Name: _____

District: _____

School: _____

Address: _____

_____

Phone number
(Indicate area code) _____

Please mail the completed form to:

Division of Educational Testing
Room 771 EBA
New York State Education Department
Albany, New York 12234